Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 103-112.

Printed in Granada, Spain.

GUDMUND R. IVERSEN

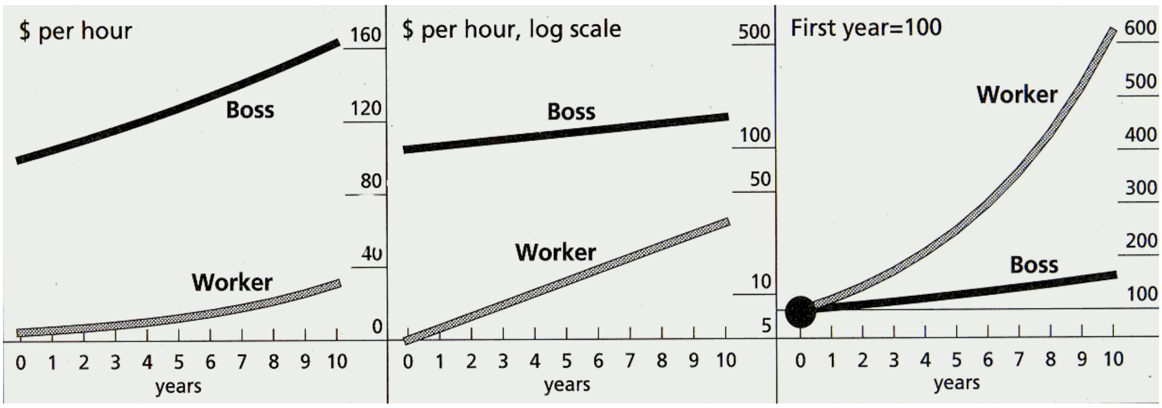
BAYESIAN MODELS AND WORLD CONTRUCTS

Statistical methods have an impact on the results of any statistical study. We do not always realise that the statistical methods act in such a way as to create a construction of the world. We should therefore be more aware of the role of statistics in research, and the question is not so much about what we teach researchers but that we train them to be aware of the impact of the methods they use. This becomes particularly important in statistical inference where we have the choice between the classical, frequentist approach and the Bayesian approach. The two approaches create very different views of the world. The paper explores the relationship between a model chosen before the analysis and the construction of the world after the analysis. Bayesian statistics may ease the conflict between model and construct.

1. WHO IS BETTER OFF?

The English weekly newsmagazine *The Economist* once showed Figure 1 in an article written as part of a series on statistics (Source: *The Economist*, May 16, 1998, p. 79).

Figure 1. A Comparison of Wages for Bosses and Workers



The purpose of the graph was to make a comparison between the wages of bosses and workers. The comparison was made with time series data over a ten-year span, and the graphs plot three aspects of wages against time.

2. COMPARING GROUPS

Statisticians are very good at comparing groups. Typically, the comparison of two groups is made into a comparison of two means or, perhaps, a comparison of two percentages. If necessary, it is even possible to compare quantities such as two variances or two regression coefficients. Statisticians are good at computing the proper test statistic and finding the resulting *p*-value that will help decide whether the difference between the groups is statistically significant or not.

But are statisticians doing the "right" thing by making such comparisons? What does it mean to say that two groups are different with respect to some variable? A researcher brings a statistician data on two interval/ratio variables for observations in two groups and asks for help to find out whether the groups are different or not. The researcher would typically have had a statistics course or two, particularly if she is doing biological research. If she is a social scientist, her training in statistics may have been rather minimal. The statistician's immediate response is to do a *t*-test for the difference between two means, assuming no wild departures from normality and that the underlying variances are not strikingly different. The statistician enters the data in a statistical software package of some kind and asks for the *t*-test. Based on the *p*-value returned by the software package, the statistician will tell the researchers whether there is a statistically significant difference between the two groups or not.

Statisticians are conditioned to do this from a long career in statistics, and hardly ever do they stop and consider whether they have done the right thing. Does it make any sense to compare the means? The statistician, forcing the comparison of two groups into a comparison of two means, implies that the statistician has constructed a reality of the world for the researcher, whether she wanted it or not. Maybe the comparison of means forces the re-searchers down paths they have not intended to take. After all, there are many ways in which things can be different.

Two groups being different can mean that *all* the observed values in one group are larger than the observed values in the other group. "Different" can mean that *some* of the values in one group are larger than *some* of the values in the second group. Different can mean that some type of an *average* is larger in one group than the other group, be it the mean, the median, or whatever the favourite average may turn out to be. So, what is the meaning of "different" when it comes to a comparison between Workers and bosses? What can we conclude about what that world out there *really* is like for the two groups?

The graphs in Figure 1 provide quite an eye-opener for students when they are exposed to them. First, it is possible to probe the students to see what their thoughts are on the meaning of the statement that two groups are different. The end of the class discussion usually consists of an agreement that two groups are different on some variable if the two corresponding means are different. Students sometimes go as far as saying that it could be that the two medians are different, when they remember something about skewed distributions. When the question is made more concrete, and students are asked to think of a comparison between wages for blue-collar worker and white-collar workers, they usually respond that the white-collar workers would be expected to have higher wages, and so the groups are different.

When students are pushed, they propose that there is a list somewhere containing wages, and they base their answer on the existence of such data. The implication of their answer is that they base their thinking on the existence of a true fact out there, in the "real world." There are wages out there, and different groups have different wages. That is a fact about the world.

The students think they know what they mean by the study of the difference between two groups, but do they really? Are there *facts* out there, waiting for statistics to be discovered? The same discussion in a group of statisticians would not have been very different; perhaps more sophisticated, but in the end not very different.

However, when faced with the graphs about wages for workers and bosses, students are no longer as certain about the factual world as they were in the beginning. The figure contains three different graphs displaying the same factual world, but the conclusions from the three graphs are very different. According to the first graph, the wages in dollars per hour are plotted as a dependent variable on the vertical axis against time as an independent variable on the horizontal axis. For ease of comparisons, the points for each group have been connected by curves, which results in two curves, one for the bosses and one for the workers. The top hourly pay is $160 for the bosses in the last year.

The first graph shows the curve for the bosses to be considerably higher than the curve for the workers across the ten years. From that it may be possible to conclude that the bosses are better off than the workers are. Is that is the way the world really is? Is that a fact that has been uncovered about the world? Or is this maybe simply a construction of the world we have created and are now, as statisticians, forcing on the researcher and thereby on those who read the research report? Does having more money even mean being better off? And does the researcher recognise that we, the statisticians, have added something to the research finding? The result is not just the data speaking, it is also a particular way of displaying the data that is speaking.

Turning to the middle graph, the dependent variable has been changed. Instead of actual wages per hour, the wages have been transformed into logarithms. This bothers students right away. They say they do not live on logarithms of money, they spend real dollars and cents. Students are not as familiar with logarithms as they used to be, now that cheap calculators are readily available for multiplication and divisions. After some discussion, however, it is possible to get students to understand that the use logarithms of wages makes it possible to see percentage increases over time, and that is what the middle chart shows.

In the middle chart, the points for the years are again connected to give us two curves, one for the workers and one for the bosses. The curve for the bosses still lies above the curve for the workers, but now the curve for the workers is rising faster than the curve for the bosses. Somehow, the workers are gaining on the bosses. Maybe they will even pass the bosses some day. The workers may not be as badly off, after all. Well, this is a different reality we use statistics for to construct and paint for the researcher. What is the researcher to do? All of a sudden, maybe statistics is not as helpful as she thought it would be.

The third graph, on the right, again shows two curves. But this time the curve for the workers lies above the curve for the bosses. Earlier, we thought we had shown that the bosses are better off than the workers are! The graph shows annual wages, and setting both wages equal to 100 at the beginning of the time period compares them over time from a common base. The curve for the workers now goes up much more rapidly than the one for the bosses, meaning that the workers have gained more than the bosses have.

3. OUR CONSTRUCTION OF THE WORLD

Each one of these three graphs paints a different picture of the world. The same data are used in the three situations. Still, through statistics, we have constructed three different realities. So, which is the "right picture"? The obvious answer is that none of them is the right picture; it all depends upon how we look at it. This may not be what the researcher wants to learn. And this may be where we have not taught our students, turning into researchers, that *the results all depend.*

Lawyers are used to thinking this way. A case before a judge and a jury is not so much about what is true and false, and whether there is a real world out there? Instead, it is getting a client acquitted, in case of the defence, and getting the accused to be found guilty, in the case of the prosecutor. From time to time both sides even invite statisticians to appear as expert witnesses. It is always amazing to see how two statisticians can use the very same data to arrive at very different conclusions. Statisticians on the two sides of the case construct two very different realities of what the world is like and hope that the jury will accept their constructs.

In answer to the question of what we should teach researchers, maybe we should include in our courses a visit to a courtroom and watch statisticians in action. That will very quickly make anyone realise that there is not one, factual world, for us to discover. It is not that researchers should learn certain, specific statistical methods, and after they learn those methods, all is well. Instead, it all depends. Maybe that is the one thing researchers should learn from us: The results heavily depends on the statistical method used, not just the data, and we should not worry about whether they know time series analysis or incomplete two-way analysis of variance or whatever else we teach. Maybe we should not even teach them statistics; they should come to statisticians for their statistical analyses.

4. LIBERAL ARTS EDUCATION

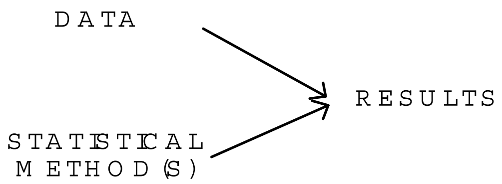
At this point, let us take a small detour into the American system of higher education. Perhaps the greatest American contribution to higher learning consists of the system of liberal arts education. For four years, after secondary school, students attend a liberal arts curriculum, are not expected to learn a profession, but instead are expected to immerse themselves in a liberal arts way of approaching life. Those who teach at such a liberal arts college, their task is not so much to teach specific statistical methods as it is to convey a way of statistical thinking to the students, based on randomness, variation and statistical regularities. The hope is that such an approach will make the students better citizens. If they need technical training in statistics for a career, they will get that through their graduate studies after college.

5. EFFECTS OF DATA AND OF METHOD

Schematically, the discussion above can be illustrated as shown in Figure 2. Thus, to consider the training of researchers in statistics, we should consider the schematic view of the research process as it is shown in the graph in Figure 2.

The graph picks up on how the results obtained from a particular research project come from two sources. One source is the data, obviously, and the other source is the statistical method. This little figure always surprises people. We like to think that results from a research process somehow are “The Truth” about the topic being studied. The purpose of teaching students statistics in a liberal arts program is not to make them into researchers or amateur statisticians, being able to do empirical research themselves. Instead, the purpose is for them to be able to understand the role played by statistics in today's society. Hopefully, this graph helps them better to understand the role of statistics in research. Having seen the three sets of curves from *The Economist*, they begin to appreciate the role played by statistical methods, beyond the data themselves.

Figure 2. Schematic View of the Research Process



We are used to thinking that the data affects the results, and we teach about the presence of sampling variation. This aspect of the research makes sense to the students. What they find discouraging and surprising is that the statistical methods themselves somehow can have an effect on the results. When they learn that *the* correlation between two variables equals 0.87, they like to take this as a fact about the relationship between the two variables in the same way as, in physics, a metal has a specific gravity constant. The students get disappointed when they learn that this correlation is the number we get when we base the analysis on least squares. For example, had we used absolute values instead to fit a line, then the result would have been different. Any other measure of the strength of the relationship between two variables is similarly dependent on how it is defined. Again, the impact of the method shows its ugly presence.

So, how are we to look at the end result of a research process? What should we, as statisticians, impress upon researchers who make use of our methods? We have the responsibility to stress that any statistical result from a research process represents a construction of the world created jointly by our data and by our methods. Just as with the three graphs from *The Economist*, there is no Truth out in the world with a capital T. The results obtained from the empirical world consist of a *construct of the world* the researchers create. It seems as if statisticians often forget that. We gather to discuss what researchers should know about statistics when they go about their tasks. We can make up a wish list of statistical methods that it would be nice if researchers knew how to use in their work. But our work is not done by simply producing such a list. Statisticians have not executed their responsibilities if that is all they do.

6. RESULTS OF THE RESEARCH PROCESS

Statisticians need to do more. They need to make researchers aware of the fact that the result of a research process is a particular construction of the world. This construction comes from the combination of our data and our statistical methods. One implication of this is that it is not as important *what* we teach researchers as it is important that they recognise the full implication of using statistics.

It is tempting to think as a parent and ask what tools do we let our children work with as they grow up. We hesitate to let a young child play with a chain saw, and perhaps we should hesitate letting researchers have access to certain statistical methods unless they are fully prepared and ready for such usage.

One good thing about the pre-computer world of statistics was that statistical methods were not as easily accessible as they are today. It used to be, that to do a ten-variable regression analysis, one put in a good bit of thought about whether it was worth doing before employing several graduate students to compute sums of squares and sums of cross products and invert matrices.

Now, this question does not even come up any more about whether a particular analysis is worth doing or not. With a few clicks of a computer mouse, the results are there for all to see in a matter of a very short time. Maybe this is not necessarily such a good thing. Maybe statisticians should require people to have a license before they are allowed to use a method such as multiple regression. It is scary to think of all the many misuses that have taken place with such analyses, because the tool is so readily available. In the wrong hands, multiple regression software may be as dangerous as the chain saw in the wrong hands.

Clearly, researchers should know multiple regression, and clearly we should encourage researchers to use multiple regression. But, perhaps we forget to let people know what they are getting into. It has already been alluded to the fact that an analysis based on least squares would give results different from an analysis based on absolute values. Just because historically it was computationally more appealing to use squares than absolute values, and the derivative of a square is easier to work with than the derivative of the absolute value function, we should not necessarily continue to use squares. But more than that, it is important that we tell researchers what they are getting into by using something like regression analysis.

Statisticians do not necessarily tell researchers that their results are partly a consequence of the methods they have been taught. Statisticians let people think that the sums of squares they get are *the* measures of the impacts of different variables. We also let people think that *the* p-value they get for a coefficient is a measure of the degree to which we can reject a null hypothesis and assess the impact of a particular variable.

7. DESCRIPTIVE AND INFERENTIAL STATISTICS

All empirical fields need some way of simplifying their data, and let us distinguish for a moment between descriptive and inferential statistics. Descriptive statistics/exploratory data analysis is not as controversial as inferential statistics, even though both impose something on the research process, and this is not always fully acknowledged. The impact of descriptive statistics is less controversial, and below the discussion is limited to the impact of inferential statistics. Here, of course, the major choice is between classical and Bayesian statistics. More specifically, through the uses of inferential statistics, statisticians impose a major construction of the world based on the results of empirical research.

Statistics is a funny field. Statisticians teach it, but mathematicians, economists, biologists and many other folks also teach it. Statisticians have completely lost control of their field, if they ever had such control, and they have let other people step into the vacuum left by statisticians. One can wonder what a sociology department would say if a department of mathematics and statistics started teaching sociology just because some of the faculty had a couple of courses in sociology in years past.

The goal of most statistics training seems to be to make the students into amateur statisticians who can tell when to use a t-test and when to use a chi-square test, but not have any understanding of why they are doing what they are doing. It would seem that the danger of doing this is so great that perhaps we should not teach statistics to other than statisticians. Through our inferential methods we are imposing a worldview that the world perhaps is not ready for.

8. STUDENT REACTION TO INFERENTIAL STATISTICS

The two competing views, classical versus Bayesian, are very different in their approaches to statistical inference. Most discussions of which of the two to use have centred on preferences expressed by statisticians. Statisticians have not listened to their customers and heard anything about what view of the world the customers think should be imposed. Do statisticians realise the heavy hand imposed by statistics, and even more importantly, do the users of statistics realise those heavy hands? And how is it that statisticians have the right to tell the customers what is best for them? If statistics is to be taught, perhaps statisticians should listen more to their customers, the users of statistics.

At ICOTS V in Singapore in 1998, I gave a talk on some of these issues (Iversen, 1988). I said, here I am, a most excellent teacher of statistics who can explain the most obscure points with great clarity. I tell my students there will be a question about the interpretation of a specific confidence interval on the next test, and then I get these kinds of answers:

"95% of the intervals would fall between the two values of the parameter."
"95% of the intervals will lie in this interval."
"95% chance that the actual value will be contained within the confidence interval."
"95% sure that μ is between the two numbers."
"95% of the data will fall within this interval."

These, and other answers, probably show that the students have not studied the right part of the book, and the instructor has not reached them in the treatment of the material in class, particularly if they did not attend class that crucial day when confidence intervals were introduced. But more than that, I see these answers as cries in the wilderness about how the world view we try to construct for our customers is not a world view our customers are comfortable with.

9. THE MANY MEANINGS OF THE WORD *PROBABILITY*.

Maybe one reason why statisticians have such difficulties teaching frequentist statistical inference, aside from the fact that it goes about drawing conclusions in strange ways, is the insistence on the use of the word *probability*. That word carries with it the notion of uncertainty, and that is the reason why many students have difficulties with the concept of a p-value. When the p-value is introduced for the first time and defined as the probability of rejecting a true null hypothesis, such a definition is received by a sea of blank stares. What we are uncertain about is whether the null hypothesis is true or not, and so here seems to be a way of deciding about that. As we know, this is not so.

One way around this difficulty and help students create the worldview statisticians have in mind, is to limit the usage of the word probability and ask "how often?" instead. When the p-value is described as a proportion which tells us how often we get the observed or more extreme data from a population where the null hypothesis is true, then students can construct the world view statisticians have in mind. Statisticians cannot blame researchers for constructing a random view of the world since the probability word is so misunderstood that way.

Students have no difficulties asking what is the probability that a population mean μ

is larger than 100, say. They ask this question right after we start in on statistical inference and long before any mention has been made of Bayesian statistics. Our job is to teach them that the question makes no sense. Instead, what if we ask *how often* is μ larger than 100? Then students answer always or never, depending on the value of μ. Maybe statisticians should banish the use of the world *probability* and substitute *how often*, instead, if we stay with the frequentist approach. Then, perhaps we can stay frequentists and still be honest with ourselves.

10. ROLE OF THE STATISTICAL MODEL

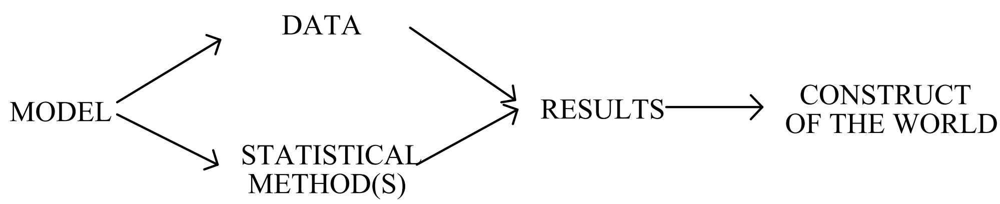
What has been discussed so far is obviously the role of the statistical model in the research process, even though it has not been explicitly said so yet. We were all taught that the model should be stated in terms of properties of the world we wanted to study. From the model we would then come up with the proper statistical method and guidance on what types of data should be gathered. Finally, we would examine in what way the data fit the model we had prescribed. As a particularly beautiful example, we can make a few assumptions of how telephone calls could be expected to arrive, and from that it is possible to derive the Poisson distribution.

How does the issue of a model relate to the construction of the world we come up with after the data have been collected and analysed? I will argue that for many users of statistics, there is a big difference between the model being used and the construction of the world that arises as a consequence of the statistical analysis by the researcher. We may think that the two are the same, and we may want the two to be the same, but often they are not. It is in the difference between the two we may find clues to how we should teach statistics to researchers, particularly how to teach statistical inference, in such a way that they understand what those ideas are all about.

Figure 3 shows an expanded view of the earlier Figure 2 on the research process. On the left side the term "Model" has been added, and on the right side the term "Construct of the world" have been added. The arrows indicate how we start with a model, which then dictates the choice of data and statistical method. From those come the results, and finally the results of the analysis are used to construct a worldview.

How the model and the construct relate to each other? Are they the same or are they different? Ideally, the model arises from the substantive issue at hand, and the construct of the world is the model in its updated form. If the model is one of the standard statistical models we use and it is not fully understand by the researchers, then there is a risk that the construct and model are in conflict with each other.

Figure 3. Schematic View of the Research Process with Model and World Construct



In its most elementary form, a model is something that simply stands for something else. It can be a physical item, such as a model toy car, or it can be a model expressed in some mathematical way. We like to think that the model represents the phenomenon we are studying, and by gathering and analysing data, we will learn more about the model and thereby the world that it stands for. Statisticians are used to thinking that way, and because they understand the model, their construction of the world will resemble the model. We understand that finding a correlation coefficient of 0.87 implies a model of linear, least square analysis, with its strengths and weaknesses. But not all researchers using statistics think this way. The users often construct their worldview in such a way that the 0.87 becomes a property of the world, as they see it.

How do non-statisticians select and use statistical models? This raises the question of the origins of our models in relation to the uses of our statistical methods. It seems as if many researchers use statistical methods with scant thought to why or why not they should use a certain method. Methods are most often used for their convenience and not because the underlying model in some way fits the problem at hand. The choice of model may not be the correct one, and the data may not even satisfy all the requisite assumptions needed for the use of the model.

All this is particularly true in statistical inference. The two major competitors represent very different models, and they construct their own views of reality. The frequentist and Bayesian constructs are very different, and how does a researcher make a choice between the two? How does a researcher chose between the frequentist and the Bayesian construction of reality? Backing away from that question for a minute, how do researchers view and understand the reality constructed by each of these two approaches?

The attempts of the students to express the idea of a confidence interval in ways that make sense to them, shows a major display of the types of difficulties we are facing in our teaching of statistics to users of statistics. It seems that, for better or for worse, if we teach the students statistics, we should teach them methods such that their construct of the world is as close to the model as possible.

This argument takes us directly to Bayesian statistics. The quotes above on confidence intervals show that many students unknowingly introduce Bayesian interpretations even when they try to do classical inference. In the Bayesian realm, the model views the world as a random world, full of uncertainties. Students intuitively conclude that the best way to model this uncertainty is to use probabilities for what we are uncertain about, thereby taking them into the Bayesian model with prior and posterior distributions.

11. BRINGING MODEL AND WORLD CONSTRUCT CLOSER

So, where does all this leave us? I have argued that researchers construct their view of reality as a function of the data and the statistical method they use. They often overlook or misunderstand the role of the statistical method, and the statistical model that gave rise to the particular method does not necessarily agree with the researchers' construction of the world. This is particularly so in statistical inference. There, to bring the model and construct together, we owe it to the world to use Bayesian statistics. That way, we are permitted to deal with the uncertainty we have about population parameters, and this is the way many researchers construct their worldview.

This is particularly so for people with only a weak background in statistics. But even well trained researchers, who use frequentist methods, very often interpret their results in a Bayesian probabilistic way. Even for many experienced researchers, when they explain what a researcher has learned after rejecting a null hypothesis, we would find traces of Bayesianism. As long as we are uncertain about values of parameters, we will fall into the Bayesian camp. As statisticians, we owe it to researchers using statistics in their research to make clear the impact statistics has on their work and enable them to choose Bayesian methods. We should train researchers well enough to make it possible for them to understand the role Bayesian statistics can play in their work. That way, the worldview they construct may actually be a reflection of their models.

ACKNOWLEDGMENTS

I am very grateful to Mary Gergen, Roberta Rehner Iversen, George Cobb, Thomas Moore, Rosemary Roberts, Milo Schield and Jeffrey Witmer for comments on earlier drafts of this paper. I also want to acknowledge my intellectual debt to L. Jimmie Savage. Working for him as a research assistant was a unique and very rewarding experience.

REFERENCES

Iversen, G. R. (1998). Student perceptions of Bayesian statistics. In L. Pereira-Mendoza, L. Seu Kea, , T. Wee Kee, & W. K. Wong (Eds.) (1998), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 231-238). Singapore: International Association for Statistical Education and International Statistics Institute.

Savage, L. J. (1981). *The Writings of Leonard Jimmie Savage. A Memorial Selection*, Washington D.C.: The American Statistical Association and The Institute of Mathematical Statistics.

Gudmund Iversen
Department of Mathematics and Statistics,
Swarthmore College, 500 College Avenue
Swarthmore, PA 19081, USA.
E-mail: iversen@swarthmore.edu